

5. Долгих И. В. Совершенствование методов оценивания сущности брендов с использованием факторного анализа // Материалы 54 международной научной студенческой конференции (МНСК-2016), 16–20 апр. 2016 г. – Новосибирск: Изд-во НГТУ, 2016. – С. 20-22 - ISBN 978-5-4437-0498-2

6. Эконометрика: учеб.-метод. комплекс для дистанц. Обучения / сост.: А. Л. Осипов, В. Н. Храпов. – Новосибирск: Изд-во СибАГС, 2002. – 172 с.

7. Joshua B. Tenenbaum, Vin de Silva, John C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science Vol 290, 22 December 2000, 2319-2323.

## ПРОБЛЕМА КАЧЕСТВА ИСХОДНЫХ ДАННЫХ В МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ В РФ

*С. В. Романчуков*

*(г. Томск, Томский политехнический университет)*

*ino@vtomske.ru*

## PROBLEM OF THE INITIAL DATA QUALITY IN THE MATHEMATICAL MODELING OF SOCIAL AND ECONOMIC PROCESSES IN THE RUSSIAN FEDERATION

*Sergey Romanchukov*

*(Tomsk, Tomsk Polytechnic University)*

**Abstract.** This article is devoted to the problems of the initial data quality in the model development of socio-economic processes in Russia. It addresses to such weaknesses in modern Russian sociology and comparativistics as insufficient degree of automation and research software support, low degree of standardization of research procedures and incompatibility of obtained results from different regions. The main task of this article is to outline the boundaries of the problem and re-energize the process of its discussion.

**Keywords:** social studies, data analysis, region, models, indexes.

**Введение.** В текущем состоянии российской и мировой экономики и социальной сферы особую важность приобретают исследования, позволяющие оценить динамику развития регионов, построить описательные и прогностические модели, позволяющие повысить эффективность управления за счёт повышения точности прогнозирования происходящих социально-экономических процессов и возможных последствий запланированных мероприятий. Значимость изысканий в данной сфере подтверждается так же количеством осуществляемых на территории РФ проектов по региональной компаративистике, однако, зачастую подобные проекты оказываются ограничены рамками сугубо социологии и математической статистики, в то время как более активное применение математических методов, информационных и сетевых технологий, методик вычислительного эксперимента ограничено в силу целого ряда проблем. Результаты применения таких методов крайне зависимы от качества, объёма, достоверности, доступности, полноты и степени формализации исходных данных.

**Проблема недостаточной информатизации социальных исследований.** Разумеется, данная проблема актуальна не первый год, и по этой теме опубликовано достаточно большое количество материалов, однако даже их авторы зачастую отмечают тот факт, что разработанные в рамках информационных технологий компьютерные приёмы решения социологических задач, остающиеся неизвестными большинству социологов и психологов, "указанный автор (прим. Давыдов А.А.) долго и тщетно пытался привлечь внимание к ним социологов"[1] и потому зачастую не внедряются в реальную каждодневную практику исследовательских групп. Кроме того акценты в большинстве подобных публикаций смещены в сторону задач анализа и обработки социологических данных (как например работы Толстой Ю. Н., Давыдова А.А. и других), которые является важными и значимыми, но отнюдь не

единственными проблемами при организации социального исследования. Разработка информационных систем, обеспечивающих обмен данными между исследовательскими коллективами, таких, например, как Информационная система "Модернизация" ЦИСИ ИФРАН,[2] является даже для крупных коллективов одной из побочных задач и не ведёт к качественному улучшению ситуации.

Одним из последствий низкой вовлечённости информационных технологий в процесс социального исследования оказывается низкая степень доступности его результатов, даже в тех случаях, когда они предполагаются как общедоступные. Зачастую, даже при наличии специализированных информационных систем, предназначенных для обмена результатами работы и исходными данными, получить актуальную информацию можно лишь при личном контакте с участниками проекта, так как наполнение и обновление баз данных таких систем происходит с многолетним запозданием.

Не менее важным следствием данной проблемы становится оперативность процессов сбора данных и обмена ими как внутри исследовательского коллектива, так и тем более между коллективами. Нередки случаи, когда результаты социологического опроса, проведённого в одном году, оказываются проверены и полностью переведены в формат, пригодный для дальнейшей обработки лишь в следующем году, а реальные исследования и публикация научных работ по ним начинается ещё два-три года спустя, что с учётом динамики современных социально-экономических процессов абсолютно недопустимо.

**Проблема сопоставимости данных.** Не менее важная проблема при попытках использовать в построении математической модели того или иного социального процесса, наблюдения которого проводились в разных регионах - проблема сопоставимости полученных результатов. Даже такие банальные и базовые индикаторы, как, например, Индекс социального самочувствия, вычисляемый на основе данных массового опроса населения, как среднее трех субиндексов: коэффициента защищенности населения, степени удовлетворенности населения своей жизнью в целом, среднего значения социального оптимизма, при всей своей видимой простоте, вычисляется в различных коллективах и проектах на основе различных наборов исходных данных с использованием зачастую абсолютно произвольных вычислительных процедур. В результате целесообразность использования даже индексов и индикаторов, носящих одинаковое название, но вычисленных в рамках разных федеральных проектов остаётся сомнительной.

Данная проблема, разумеется, также не является секретом для социологов, и в документации любого крупного федерального исследовательского проекта (например, [3]) ей и методам её решения уделяется особое значение. Даже на уровне одного проекта, имеющего несколько подразделений, необходимость обеспечения сопоставимости результатов доставляет серьёзные неудобства, но при попытках использовать материалы от нескольких разных проектов в качестве исходных данных в вычислениях и при построении математической модели влияние этого фактора усиливается многократно.

**Размышление о преодолении последствий.** На тактическом уровне, при построении математической модели конкретного социального процесса, можно попытаться уйти от данной проблемы, но ценой повышения трудоёмкости исследования:

Во-первых, при выборе исходных данных предпочтение необходимо отдавать "сырым" данным, нежели рассчитанным на их основе индексам и показателям. Так, например, если модель предполагает использование ИСС как одной из входных переменных, то вычислав его самостоятельно для наборов данных, полученных от разных коллективов, можно быть более уверенным, в сопоставимости полученных результатов, нежели получив от тех же коллективов сразу числовые значения индекса.

Во-вторых, как показывает практика, даже при работе с сырыми данными крупных проектов необходимо перепроверять их достоверность (репрезентативность выборок и т.д.). Ни один источник социологических данных, кроме, возможно, Росстата, нельзя считать априори надёжным. Здесь следует помнить также о том, что и технологии контроля качества

работы интервьюеров (как низового звена, собирающего исходные данные в ходе социологических опросов) на текущем этапе нуждается в улучшении [4].

В третьих, невозможно недооценить важность и необходимость поддержания личных контактов со всеми исследовательскими коллективами, ведущими работы в области, даже в малейшей степени смежной с областью проводимого исследования.

В общем же размышления о данной проблеме приводят нас к дискуссии о необходимости переустройства отечественной социологической науки в новом тысячелетии и о степени соответствия используемых в ней информационных сетей и систем потребностям времени.

**Заключение.** Данное сообщение обращено прежде всего к студентам, планирующим развиваться в области математического и информационного сопровождения социальных исследований и математического моделирования масштабных социальных процессов. Проблемы, затронутые статье существуют уже давно, и неоднократно обсуждались, но до сих пор не решены, поэтому как минимум их обсуждение остаётся по-прежнему актуальным.

## ЛИТЕРАТУРА

1. Толстова Ю. Н. Социология и компьютерные технологии / Социологические исследования, № 8, Август 2015, С. 3-13
2. Информационная система "Модернизация" ЦИСИ ИФРАН [Электронный ресурс] URL: <http://mod.vscs.ac.ru/> требуется авторизация
3. Лапин Н.И., Беляева Л.А. Программа и типовой инструментарий «Социокультурный портрет региона России» (Модификация – 2010). М.: ИФРАН, 2010
4. Романчуков С.В., Берестнева Е.В., Маклакова Т.Г., Шухарев С.О., Информационная технология оценки качества работы интервьюеров. Труды Конгресса по интеллектуальным системам и информационным технологиям (IS-IT' 16) - Дивногорское, 2-9 сентября 2016. – Таганрог: ЮФУ, 2016. – Т. 1 - С. 275-278

## ОШИБКИ И НЕДОСТАТКИ СИНТАКСИЧЕСКОГО АНАЛИЗА СЕМАНТИКО-СИНТАКСИЧЕСКОГО АНАЛИЗАТОРА ТЕКСТА SEMSIN

*В. В. Чемерилов, А.С. Фадеев.  
(г. Томск, Томский политехнический университет)  
e-mail: vchemerilov@gmail.com*

## THE SYNTAX ERRORS AND DISADVANTAGES OF THE SEMANTICS AND SYNTAX ANALYSIS OF THE SEMSIN PARSER.

*V.V. Chemerilov, A.S. Fadeev  
(Tomsk, Tomsk Polytechnic University)*

**Annotation.** This article describes algorithm of the Semsin parser. During the research of this tool for its adaptation to the text-to-speech system, errors in syntactic analysis were found.

**Keywords:** text-to-speech systems, text parser, semantic text analysis, syntax text analysis.

**Введение.** В процессе компьютерного анализа текстовых данных, текст проходит через несколько этапов обработки. На одном из таких этапов происходит синтаксический анализ, задача которого состоит в определении синтаксических связей между словами предложения. Результаты синтаксического анализа могут быть использованы для различных целей, например, для автоматического разрешения омонимии [1]. В данной работе рассмотрены ошибки, совершаемые семантико-синтаксическим анализатором Semsin при синтаксическом анализе предложений.